

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

Appendix 2

The Narrative Report

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

ADMINISTRATIVE-INTERNAL USE ONLY

The Narrative Report

When a person outside of the Washington area completes a formal application for an Agency position, he or she is assigned to one of the field test settings to take PATB I which consists of five cognitive tests and the Strong-Campbell Interest Inventory. The answer sheets for these candidates are mailed to Washington and delivered to the office of PSS. Candidates from the Washington area are given PATB I and PATB II and their answer sheets are also sent to PSS. Answer sheets for both groups of candidates are scored but nothing more is done with them and no one sees them unless a unit of the Agency requests a write-up, i.e., a report of performance on the tests. This report is prepared by psychologists in OMS/PSS who will not give actual test scores but only their interpretation of them.

If the request for a write-up is general, that is, no specific component or job within the Agency is identified, then the write-up tends to be non-specific. If a component of the Agency or a specific job is identified, then the write-up is ^{? S/ur ?} supposed to be focused on specific job profiles for that component. We encountered the terms

*1-2. descriptive -
which is what it
should be*

ADMINISTRATIVE-INTERNAL USE ONLY

"job profiles" and "job-group profiles" a number of times in memoranda reporting interviews with the Chief of PSS or in memoranda written by the Chief of PSS. To us, these terms meant that a systematic detailed analysis of professional jobs had been done to determine the knowledges, skills and competencies needed to perform the professional jobs satisfactorily. Further investigation on our part proved this interpretation to be incorrect. No systematic job analyses have been done. To the psychological staff the terms mean test profiles that have been generated for a number of job groups in the Agency.

Although we tried to find out how these test profiles were generated, we were unable to do so to our satisfaction. In a memorandum written by the Chief of PSS to the DDA (25 July 1979), the Chief states that test profiles for a number of jobs in the Agency were generated as part of the initial development of the PATB. However, when we asked him questions about this and other aspects of the initial development of the battery, he stated that there was no material available on the early history of the development of the test except that contained in Test Data Book, No. 75, dated 1 July 1958. No test profiles are included in this

ADMINISTRATIVE-INTERNAL USE ONLY

This isn't true at all. Just because we haven't used the Marguerite + McCormick (Purdue) paradigm doesn't mean that "job analyses" have not been done.

disagree!
With the amount of interaction we have with Agency components, we have a very good idea of what various jobs require

see above ↑

many of the early test print-outs have likelihood for success in various jobs included.

~~ADMINISTRATIVE INTERNAL USE ONLY~~

from what the hell is the reference to the Test Data Book #75 dated 1 July, 1958 on the previous page?

source for any Agency jobs. In a memorandum reporting a briefing for DDA done by the C/PSS and one of his staff members, we noticed that he had stated each psychologist had a test data book to assist him or her in evaluating test results. We asked to see the books, but the C/PSS told us

that no such books existed. According to him, each new psychologist is trained by an experienced psychologist who has all of these data in his head.

not quite right. They are coached & tutored on the battery but the data is not contained solely in one person's head.

We were permitted to see, but not to examine closely because of security reasons, a sample computer print-out of the test profile for one applicant. As the C/PSS explained the print-out to us, the test profiles and job profiles appeared to be generated from the studies that had been done on PATB over its 20 years of use. We have reviewed those studies in Appendix 1 and have concluded on the basis of that review that there is no consistent or convincing evidence for the job-related validity of PATB. We also

negligence of the first magnitude. They never asked to sit down w/ staff psychologists to see how reports are generated from test print-outs.

Nonsense.

we don't generate data here, we report what we have.

pointed out in that Appendix that all of the studies used small samples composed largely of white males and only two of the studies had been cross-validated. In neither of the cross-validation studies were the findings of the first study verified, which indicates that the job-related validity

Absolutely incorrect. My AARP study clearly does exactly the opposite.

~~ADMINISTRATIVE INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

of PATB still needs to be demonstrated. Any test profiles for specific jobs in the Agency that were generated from these sources would be unreliable because of the small samples used, of extremely doubtful validity, and probably biased against minorities and women because these are underrepresented in the samples used in the studies. ?

*Unprofessional
and unwarranted
conclusion.*

Since we are not absolutely sure that the test profiles and job profiles have been generated from the studies that have been done on PATB over the past 20 years, let's assume that they were generated at the time of the initial construction of PATB in the 1950's by testing personnel in professional jobs at that time. Would such test profiles constitute evidence for the validity of PATB? The answer is no. The fact that a group of current employees had a particular test profile is merely description. It does not provide the evidence needed to determine whether applicants for the same positions must have the same test profile to perform satisfactorily on the job. As a matter of fact, since the test profiles for the original group represent the average score for a number of individuals, most of the individuals in the original group would not have had that profile. If the test profiles were based on an early

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

This is a good example of faulty logic. "Underrepresentation is no basis for claiming unfairness. What is underrepresentation? Maybe Blacks got a break by being 'underrepresented.' After all, it was the 'better' ones that were represented."

in the 1950's to repeat we didn't know color et al

group of employees, their use with the applicant group would be unfair to minorities, particularly, because they were underrepresented in the 1950's employed group.

Predictive Aspects of the Narrative Report

The section of the narrative report that is most directly affected by using an invalid and unreliable data base is the last one, Comments and/or Recommendations. We examined 21 "sanitized" narrative reports, 13 of which had this section completed. In 11 of the 13, the narrative report recommends the applicant for a specific job or to a specific unit of the Agency. We rechecked all of the validity data that we had and could find no evidence that would support any of these recommendations. This troubles us. Recommendations for specific types of employment made without adequate validity data promote unfair use of the test results. Such recommendations tend to lead to the exclusion from consideration for employment those individuals who score low on the cognitive tests or who have "unfavorable" scores on the other scales when there is no evidence that these people could not perform satisfactorily on the job. This practice violates EEOC guidelines on fairness as indicated in the quotation below.

→ They totally discount the fact that we probably have talked w/ mgmt and employees of the various target groups.

performing satisfactorily on a job in the Agency involves more than just performing the job duties.

ADMINISTRATIVE-INTERNAL USE ONLY

When members of one race, sex or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the groups that obtain the lower scores.^{1/}

Before ending the discussions of the Comments and/or Recommendation section of the narrative report, we think that we should make a few additional comments concerning some statements frequently made by the staff of PSS that are related to this section. In a number of reports of interviews with the Chief and staff of PSS, the C/PSS is reported as stating that no cutoff scores are used for PATB, that test results are never used in a pass/fail context, and that PSS has no role in hiring decisions. Although it is true that no single cutoff score for each test is used to screen out applicants and that the pass/fail designation

^{1/} Equal Employment Opportunity Commission. Uniform guidelines on employee selection procedures (1978). Federal Register, August 25, 1978, 43, (166), p. 38301.

is not directly used, indirectly both are used when the unvalidated equations and profiles are used to make a recommendation to hire or not to hire an applicant.

To say that PSS has no role in hiring decisions is dissemblance of the highest order. PSS, through its narrative reports, plays a significant role in some hiring decisions. From reports of interviews with people in different units of the Agency who apparently have the responsibility for making the final selection decisions, it is quite clear that a significant proportion of the decisions to hire or not to hire are greatly influenced by the narrative report, particularly the recommendations made by PSS.

From these interview reports, one would conclude that the failure of PSS to recommend an applicant is equivalent to a "kiss-of-death" for that applicant in some of the units. This makes the recommendation section of the narrative report even more troublesome because PSS makes its recommendations with a level of confidence and finality that is not supported by the validity and reliability of the data.

*a judgment which
is, in my opinion,
absolutely false*

We have recommended in Appendix I that operational use of the multiple regression and discriminant analysis equations be discontinued until the equations have been

*is this
consistent with
what they said
before*

*⇒ illiteracy is
the single,
most common
reason for
rejecting appli-
cants. It is
quite likely
that basic
literacy is
important in
most jobs.*

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

This could take years and would have to be at the initiative of the USER office, Not PSS

cross-validated. Since the recommendation section of the narrative report represents the operational use of these equations, we think that their use for this section of the report should also be discontinued. The best use of this section of the narrative report would be for summarizing descriptively the strengths and weaknesses of the applicant.

Descriptive Aspects of the Narrative Report

One, and probably the most important, function of the narrative report is to provide a clear, accurate and meaningful description of an applicant's characteristics as revealed by PATB. Persons in the various units of the Agency can then use this description together with other sources of information about the candidate such as the Personal History Statement, transcripts from educational institutions, and letters of recommendation to arrive at employment decisions. By using a variety of sources of information, persons in the units should be able to make employment decisions that are beneficial to the Agency and fair and equitable for all candidates.

This, essentially, is what happens.

true!

Six of the seven sections of the narrative report are intended to be descriptive. Two sections describe performance on the intellectual tests of PATB and one section is devoted to each of the following: (1) measured vocational

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

interests, (2) foreign language, (3) writing ability, and (4) attitudinal and personality factors. The value of these descriptive portions depends upon two major factors, the validity and reliability of the individual tests and how well written the descriptions are.

In Appendix I, we indicated that the content and construct validity of the individual tests comprising PATB has not been determined. Although this limits the value of the description, it does not make it completely useless. The content of the Vocabulary, Reading, Contemporary Affairs Test and Numerical Operations tests clearly indicates that they are appraising what their titles suggest they are appraising. The Essay test is a writing sample and directly appraises one type of writing ability. The Strong-Campbell Interest Inventory is a standardized instrument which provides validity data in its manual to establish what it is appraising. However, we cannot infer from the content of the other tests and scales what they are measuring or what the scores on them mean.

We suspect that the Figure Matrices test measures abstract reasoning because tests of this type usually do; however, one cannot establish the validity of a test just by determining its superficial similarity with other tests. There are no data to indicate what abilities the Language

*Don't that
"Good" ?*

*Come, come!
(see attached) (3/26)
PM VEFM*

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

Aptitude, Interpretive Reasoning, and Considerations tests are appraising or what is being appraised by the work attitude and temperament scales. These types of tests and scales need to have their construct validity established. Factor analytic studies would have been extremely useful in determining what these tests are measuring but, unfortunately, no such studies are available. Without construct validity data one cannot say anything about what a score means. The psychologists who write the narrative report have tried to avoid this issue, particularly in reporting performance on the cognitive tests, by just listing the test by name and giving an adjective such as average or excellent or poor to describe the performance. As a result these descriptions are atomistic and fragmentary which makes it impossible for the reader to get a clear, comprehensive picture of the cognitive competencies of an applicant.

We may not have provided them to the contractors, but they have been done, to some extent.

We have also indicated in Appendix I that the reliabilities for many of the tests of PATB are distressingly low and, as a result, the standard errors of measurement for these tests are relatively large. We found that there were { no reliability data for minorities on any of the tests and no data for women on the work attitudes scales. In the absence of such data, one should be extremely tentative in

there is very little to base this on

things to do

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6
interpreting their performances on the tests. The narrative reports that we examined did not take this into account. They described test performance with the same level of confidence for all applicants and for all tests. This is troublesome because it leads the reader, who is usually naive in testing, to ascribe a level of accuracy and finality to the performance that is not merited by the reliabilities of the tests.

Two parts of the descriptive sections of the narrative reports, measured vocational interests and writing ability, caused us considerable concern. At the present time vocational interests are appraised with the Strong-Campbell Interest Inventory. There are no Agency norms for this test and no validity studies have been done to determine whether scores on this instrument are related to job performance.

We suspect that managers in the units are not aware of this. In addition, in 11 out of 21 reports that we read, the scores on this test were misinterpreted. In reporting these scores, the psychologists used phrases such as good verbal-persuasive skills; a rugged, practical-individual; outgoing; and strong organizational and supervisory skills. The Strong-Campbell Interest Inventory appraises none of these characteristics, and the manual specifically warns users against these types of interpretations.

1/ Campbell, David P. Manual for the Strong-Campbell Interest Inventory. Stanford, California. Stanford University Press (1974) pp 17, 21, 87, 114

ADMINISTRATIVE-INTERNAL USE ONLY

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

They probably mean 1978. 1974 version has 98 pages only.

Even if this were true, which I doubt, it would be irrelevant. The SCII is not used as an instrument to predict job performance!

directly from the SCII manual 124 and 78

users

not

users

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

The Strong-Campbell Interest Inventory is an extremely complex instrument that yields 158 scores - scores on 6 General Occupational Themes, scores on 23 Basic Interest Scales, scores on 124 Occupational Scales, an academic orientation score, an introversion-extroversion score and 3 administrative index scores. It is impossible to tell from the narrative description which score or scores are being interpreted. However, it is quite clear that a number of the psychologists who are writing this part of the narrative description *deflammatory and unsupported* do not understand the instrument. For example, on one narrative report the following description was given: "Measured vocation (sic) interests are very broad, encompassing, virtually every occupational field. This type of profile suggests a highly-motivated, versatile individual, eager to enter the world of work." An individual who has a large number of high scores on the interest inventory has marked "like" to an exceptionally large number of the items on the inventory. This type of person is discussed on page 85 of the manual as follows: "There is no single characteristic descriptive of all persons with high LP's, (note: percentage of like responses), but some combination of the adjectives "enthusiastic," "curious," "shallow," "unfocused," "energetic," "manic" will fit many

not necessary?

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

of them." As one can see, the interpretation given in the manual is at variance with the PSS psychologist's interpretation.

We think that it is impossible to compress into 10 or fewer typed lines a meaningful interpretation of a complex instrument like the Strong-Campbell Interest Inventory. To try to do so tends to misinform rather than inform the reader. For this reason and also because no validity data or norms are available for the use of the instrument for selecting personnel in the Agency, we recommend that no report be made to the units of these scores.

The part of the narrative report that describes the writing ability also caused us concern. First, the writing sample has not been validated. Second, neither the reliability of the writing sample nor the reliability of scoring or judging the writing sample has been determined. Third, the wide variation in describing the candidate's writing ability indicates to us that there are no established guidelines for scoring or judging the writing sample. Fourth, for some unexplained reason, the report of writing ability includes the candidate's claimed ability which does not appear to serve any useful purpose.

B.S.

T? We are looking solely for grammar & organization

True. When there is doubt, however, we send a copy of the essay w/ the report

It is amazing to me that these school-teachers are unable to see how a comparison between claimed writing ability and an actual sample of the individual's writing would be useful.

~~ADMINISTRATIVE-INTERNAL USE ONLY~~

The actual description of writing ability is done, on the average, with 15 words and the aspects of writing ability that are commented upon vary considerably from one narrative report to another. We suspect that much of the variation in the reports of writing ability are due more to the idiosyncrasies of the psychologists writing the report than to differences in writing abilities of the applicants. This is bothersome because one purpose of the narrative report should be to supply the managers in the units with comparable data on all candidates. The descriptions being presented are not comparable; they use ambiguous terms and leave too many blanks that the managers must fill in for themselves. For example, does the phrase, not badly written, mean the same as demonstrates well-developed writing skills? If no comments are made about errors in spelling, grammar or syntax, does it mean that the candidate made no such errors or does it mean that the particular psychologist who wrote the description did not think that the errors made were worth mentioning?

see p 5
of summary
IG report

We have three major concerns about the section of the narrative report that presents the description of attitudinal and personality factors. First, the reports assume that

ADMINISTRATIVE INTERNAL USE ONLY

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

What they mean to say is that they choose to ignore the studies which do show high correlations between these scales and job performance.

the validities of the work attitude scales and the temperament scales have been established when, in fact, they have not been. The inferences in the reports that the scales measure such attributes as gregariousness, introversion, cautiousness or introspectiveness are completely unjustified. Second, the descriptions do not reflect the low reliabilities of these scales for white males, and the absence of any reliability data for the work attitude scales for females and for minorities. Third, occasionally the psychologists appear to forget that they are describing self-reports of the applicants and describe, instead, actual behavior. For example, one report states "He is an outgoing type, who eagerly takes part in planning social activities and informal gatherings." This statement describes actual behavior and the psychologist had no data on the actual behavior. Fortunately, these kinds of misstatements do not occur very frequently. *what's new*

STATINTL

In describing the performance of an applicant on the intellectual tests, the psychologists use adjectives to describe the performance and different labels to identify the tests. The Test Data Book No. 15, 1 July 1958, gives the following adjectives that were to be used to report intellectual test scores: superior, top 5%; very high, next highest 15%; high average, next 20%; average, next 20%; low average, next 20%; poor, next 15%; and very poor, lowest 5%. In the 21 narrative reports that we examined the following

Approved For Release 2002/01/25 : CIA-RDP00-01458R000100130011-6

~~ADMINISTRATIVE - INTERNAL USE ONLY~~

adjectives were used: very superior, superior, excellent, high average, above average, average, fair, weak, poor, very poor. We were not able to find a set of guidelines for translating the scores on the intellectual tests to this set of adjectives. However, it is possible to match the adjectives to the coded scores used for the tests as shown by the following: 9=very superior; 8=superior; 7=excellent; 6=high average; 5=above average; 4=average; 3=fair; 2=weak; 1=poor; and 0=very poor. If this is indeed what is being done, then the psychologists are making finer discriminations in the test scores than are justified by the reliabilities of the tests. No explanation of the meaning of the adjectives is provided on the narrative report and it is highly probable that the user of the report will misinterpret what the adjectives are supposed to represent.

How dramatic! Don't you really mean "possible"?

The label used to identify the Figure Matrices test varies in different narrative reports; sometimes it is identified as abstract reasoning, sometimes as non-verbal reasoning, and sometimes as the ability to deal with pictorial symbols. Our experience in testing indicates that a person who is naive about tests will place a higher value on the score from this test if it is labelled as abstract reasoning than he will if it is labelled non-verbal reasoning.

Interesting that it appears okay for the school teachers to use their "experience" in testing to convince their readers about this, but when PSS falls on their own "experience", it is disallowed as "unconvincing". Research to support this?

ADMINISTRATIVE - INTERNAL USE ONLY

It would be very desirable to require everyone who writes the narrative reports to use the same labels for the test. In this instance, since there is no validity evidence that demonstrates that the Figure Matrices test is indeed measuring abstract or non-verbal reasoning, it would be better to identify it simply by its title, Figure Matrices.

We noticed in the section reporting the performance on Part I intellectual tests a statement about the applicant's claim as to the percentage of the class where his or her college grades fell. We question the usefulness of this piece of information. Our experience has shown that students are quite accurate in reporting their grade point averages, but that they are much less accurate in identifying in what percentage of the class they fall. We also question the inclusion of this information because its meaning is not clear unless one knows the selectivity of the institution attended, the distribution of grades given in that institution and other factors such as whether the student worked full time while attending college. Since applicants are supposed to supply college transcripts with their Personal History Statement, managers in the units should have the actual transcript and do not need this self-report of grades.

This lets us compare PTB scores w/ college grades + therefore call inconsistencies to the attention of managers + to get them to evaluate transcripts more carefully. I have no objection, however, to having a GPA instead

The self-report is for us, not for managers.

Nonsense. They lie like hell to get hired.

This doesn't work as well as it should.

ADMINISTRATIVE - INTERNAL USE ONLY

In the 21 narrative reports that we examined, we noted that the psychologists writing the reports missed many opportunities to write an interpretation that would lead to constructive and fair use of the test results. The APA Standards^{1/} state the test user should consider alternative interpretations of a given score. Since the psychologists are the ones who are writing the narrative reports, they are the ones to whom this standard is addressed. A good example of the failure of the psychologists to follow this standard is the report given for an applicant identified as coming from a bilingual home. The report states that Abstract Reasoning is high average; Arithmetic Reasoning is average; Reading Comprehension is weak; and Vocabulary is very poor. The psychologist failed to point out that this pattern is typical for bilinguals. A true bilingual person processes verbal information much more slowly than does the monolingual person and tends to be penalized on verbal tests that are timed. The Vocabulary test is highly speeded and the other verbal tests are somewhat speeded even for monolinguals. If the applicant is truly bilingual, then his

Research?

^{1/} Standards for Educational and Psychological Tests. Washington, D.C.: American Psychological Association, 1974, p 72.

~~ADMINISTRATIVE INTERNAL USE ONLY~~

scores on the verbal tests most probably seriously underestimate his true ability. The psychologist should have pointed this out and should have advised the reader of the report to assign greater importance to other sources of information about the person's abilities than to the test scores. If the psychologists are going to do nothing more than write a somewhat stereotyped description of test performance, and this appears to be what they are doing in the sample of reports that we read, it would be better to generate the test results by computers which can do the same job much more efficiently and economically.

*I agree, for
Skills Bank writeups.
We should get to
work on this.!!*

~~ADMINISTRATIVE INTERNAL USE ONLY~~

ADMINISTRATIVE INTERNAL USE ONLY

References

Campbell, David P. Manual for the Strong-Campbell Interest Inventory T325 (Merged Form). Stanford, California. Stanford University Press 1974.

Equal Employment Opportunity Commission Uniform Guidelines on employee selection procedures (1978). Federal Register. August 25, 1978, (166), 38290-38315.

Standards for educational and psychological tests.

Washington, D.C.: American Psychological Association, 1974.

ADMINISTRATIVE INTERNAL USE ONLY